

Regner, Tobias; Riener, Gerhard

**Working Paper**

## Motivational cherry picking

Jena Economic Research Papers, No. 2011,029

**Provided in Cooperation with:**  
Max Planck Institute of Economics

*Suggested Citation:* Regner, Tobias; Riener, Gerhard (2011) : Motivational cherry picking, Jena Economic Research Papers, No. 2011,029, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:  
<https://hdl.handle.net/10419/56887>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# JENA ECONOMIC RESEARCH PAPERS



# 2011 – 029

## Motivational Cherry Picking

by

**Tobias Regner  
Gerhard Riener**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact [markus.pasche@uni-jena.de](mailto:markus.pasche@uni-jena.de).

Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

Max Planck Institute of Economics  
Kahlaische Str. 10  
D-07745 Jena  
[www.econ.mpg.de](http://www.econ.mpg.de)

© by the author.

# Motivational Cherry Picking

Tobias Regner ♠ Gerhard Riener ◇\*

♠ *Max Planck Institute of Economics, Jena, Germany*

◇ *Friedrich Schiller University and Max Planck Institute of Economics, Jena, Germany*

June 28, 2011

## Abstract

We construct a simple three person trust game with one trustor and two trustees. The trustor has the possibility to either trust both trustees or none, while the trustees make their decisions either sequentially or simultaneously, depending on the treatment. When trustees play sequentially, follower trustees who are informed about the leader's choice are significantly less kind than in the simultaneous move treatment as well as the leader trustees. These findings can not be explained by models of inequity aversion, pure guilt aversion, or conformity. Instead, follower trustees cherry pick the motivation that serves them best. When the leader trustee played unkind, they tend to conform and play unkind, too. When the leader made a kind choice, followers seem to perceive the duty of reciprocating to the trustor as already fulfilled by the leader. While guilt works well as a motivational force in a dyadic situation, it gets alleviated easily when there is someone to shift responsibility to, like the leader in our three person game.

**JEL codes:** D03, D71, C79, C92

**Keywords:** Team production, Trust, Principal Agent, Guilt, Guilt alleviation, Conformity, False consensus effect, Lab experiment, Cherry picking

---

\*Corresponding author

We are grateful to Ashok Rai, Martin Strobel and Fabian Winter for valuable comments and helpful discussions. Rico Löbel and Markus Heinemann provided excellent research assistance.

## 1 Introduction

Working together in teams often implies that actions to complete a task are taken sequentially. One person takes – or is assigned – the role of the leader, while another person is following, knowing about the action of the leader. When a principal contracts a team, the organization of the work within the team is of utmost importance. It means considerable costs to the principal, if work is structured in a way that free-riding on other team members' contributions is facilitated. In a framework with one principal and two agents we investigate how actions taken by the lead agent affect the choice of the follower agent.<sup>1</sup>

In economics, the problem of leadership and example setting in organizations has been explored only in recent years. In the seminal work of Hermalin (1998) the problem of the leader is to convince the other agents to follow her decision to invest in a costly activity, where one possible option is to try to set an example by own sacrifice. Therefore, in a situation where the follower has less knowledge than the leader on the importance of the task, the leader can credibly convince the follower. However, what happens when the information asymmetry is not severe and the incentives of the agents are not aligned with the objectives of the leader? Does having a leader setting an example has the same effects or does the leader's action crowd out effort by followers?<sup>2</sup>

Our experiment examines the role of leadership in a stylized work team by using a mini trust game with one trustor and two trustees.<sup>3</sup> The trustor can trust both trustees simultaneously and the trustees can – independently – reciprocate this trust or not. Our treatment manipulation varies the information set of the second trustee: in the *baseline treatment*, the trustees decide simultaneously, while in the *leader-follower treatment* the leader trustee takes the decision first and the follower trustee is informed about that decision before she makes hers. None of the trustees' decisions have a financial impact on the other trustee.

It is well known from Charness and Dufwenberg (2006) and similar experiments that in a two person setting guilt can be a sufficiently strong motivation for a trustee resulting in kind behavior. But what happens, when there are two trustees? In particular, what do trustees do when they decide sequentially? Does knowing that the other trustee already cooperated affect behavior? This information may reduce

---

<sup>1</sup>This problem resembles a volunteer's dilemma problem, where the volunteer makes the first move and others, rather than following the volunteer, free-ride on her efforts. However, the volunteer's problem is a sorting problem, where subjects with certain traits chose to be the leader, for altruistic or egoistic motives.

<sup>2</sup>Some evidence of the influence of leaders can be observed in leadership giving in public good and fund-raising games provide some indication: Subjects increase contributions, when leaders did so before in public good games where the interests of the leader and the followers are aligned – Potters et al. (2007) or Güth et al. (2007) – and even when the interests are independent of each other as in a charitable giving experiments (Reinstein and Riener, Forthcoming).

<sup>3</sup>Knez and Camerer (1995) already 15 years ago argued that it is necessary to study games where three players interact, however there is still strikingly little evidence in this respect. An important exception is the recent paper by Cassar and Rigdon (2010), who look at trust in a networked game. They have one trustor and two trustees or two trustors and one trustee. An important difference is that they are interested in the relative difference between the trust towards the two trustees, so the trustor can differentiate trust between the trustees.

responsibility attribution, crowding out the feeling of guilt (therefore reducing kind behavior). Guilt is alleviated when the leader trustee cooperates, and the follower realizes responsibility towards the trustor can be shifted to the leader. As a consequence the follower trustee does not cooperate, while he may well have been motivated by guilt aversion to behave kindly in a dyadic setting.<sup>4</sup>

In the *baseline treatment* around 30% of the trustees reciprocate and we find a significant correlation between second order beliefs and choices. However, we observe substantial differences in behavior in the *leader-follower treatment*. While leaders do not behave differently than in the baseline case, followers reciprocate significantly less often. Furthermore, we find that the basic mechanisms of guilt aversion (second order beliefs and the sensitivity to guilt) are important determinants of kind behavior. Interestingly, follower trustees appear to cherry pick from the two motivations (conformity, collective guilt) that explain the treatment difference. When the leader trustee has made an unkind choice, there is a strong tendency to conform, that is, play unkind as well. When the leader made a kind choice, followers seem to perceive the duty of reciprocating to the trustor as already fulfilled by the leader and play unkind. While guilt works well as a motivational force in a dyadic situation, it gets alleviated easily when there is someone to shift responsibility to, like the leader in our three person game. They seem to behave according to the motivation that serves their self-interest best.

Hence, in a team situation agents seem to be very quick in attributing responsibility (towards the principal) to another agent who already committed to reciprocate. When it is known that the lead agent did not reciprocate the principal's trust, follower agents appear to conform to the behavior of their peer as they "shirk" as well. In addition to this finding of motivational cherry picking among agents in a principal agent setting, our paper contributes two methodological aspects to the existing literature. We propose a strategy to reduce the omitted variables bias (OVB) resulting from a false consensus effect by measuring first order beliefs over other agents' behavior to reduce the resulting omitted variables bias due to potential false consensus.<sup>5</sup> Moreover, follower behavior differs under direct response and strategy method. Followers cherry pick when they are informed what the leader did. When they are asked to decide dependent on the leader's choice, followers' behavior is not different from the baseline. In addition to the hot (direct response) vs. cold (strategy method) difference found by Brosig et al. (2003), our result suggests that the strategy method may deliver different results, when the options/counterfactuals cannot easily be ranked in a linear way, as for instance the trustee's response to increasing transfers of

---

<sup>4</sup>A similar phenomenon has been described in social psychology by Latané and Darley (1968) who called it *diffusion of responsibility* in a study of helping behavior. This diffusion of responsibility can be explained by the choice of which social norm to follow which has been initially explored by Stouffer (1949) in the sociological literature on conflicting norm behavior.

<sup>5</sup>That the false consensus effect can indeed be viewed as a omitted variables problem in a regression framework is easy to see: the omitted variable (subject's belief over the state of the world and how one should behave) has indeed an influence on her behavior and it influences the view over how other subjects expect themselves to behave (the second order belief), so the conditions for the presence of a OVB are met. Controlling for this reduces the OVB. See for example Greene (1993) for a discussion of the OVB.

the trustor in a trust game (assuming the trustee wants to reciprocate).

The design of the game allows us to distinguish between two types of behavior. Pure material payoff maximization, models that take into account distributional concerns *à la* Fehr and Schmidt (1999) or Bolton and Ockenfels (2000), and guilt aversion models *à la* Charness and Dufwenberg (2006) will produce the same behavioral predictions independent of the information structure of the game. In contrast, the information structure matters when conformity is considered as a motivation or when guilt aversion is extended to a group. Conformity models (see Akerlof (1980), Bernheim (1994), or Manski (2000)) take social interactions into account and would predict that follower trustees use the choice of their leader trustee as a guideline for appropriate behavior. That is, followers would tend to play kind/unkind when their leader also played kind/unkind. If guilt aversion is applied in a collective sense, then followers would tend to be less kind, shifting responsibility, in case their leader played kind.

Shifting responsibility has previously been studied in the economics literature on experimental labor market settings. Whether a kind decision of a subject is reciprocated depends on who actually makes the decision. It matters – for example – for subjects if a human takes a deliberate choice of being kind or a random device determines the outcome (see Charness, 2000). Agents reciprocate to kind wage offers by working harder only when they were made by a human principal. A formalization of this idea has been recently brought forward by Sebald (2010). However, it is hardly known, yet more relevant for personal relations, how responsibility can be shifted to another person<sup>6</sup>, therefore undermining feelings of guilt and its behavioral consequences. This is the context our experimental setup addresses as the kind action is taken by a human being regarding to two other subjects who take their reciprocation choices either simultaneously or sequentially, a situation very common in team work.

Besides this applied, policy-relevant perspective our paper adds to the literature on the relationship between beliefs and behavior. Dufwenberg and Gneezy (2000) is an early contribution on the correlation of second order beliefs and behavior in a trust game. A recent paper by Fischbacher and Gaechter (2010) also shows how beliefs are related to the dynamics of free riding in public goods experiments. Models of guilt aversion, see for instance Charness and Dufwenberg (2006), take these beliefs as primal force that drives the behavior of subjects in this case, with the rationale that if I do not act upon these beliefs I am knowingly letting the trustor down which imposes a psychological cost on me which depends on my sensitivity to guilt. In contrast to the previously mentioned experiments Ellingsen et al. (2010) elicit first order beliefs of recipients or trustors in dictator and trust games and reveal those first order beliefs to dictators and trustees to examine their reactions. They do not find any correlation between the revealed first order beliefs and actions and conclude that the observed correlation in other studies is due to a false consensus effect. We approach this question of causality of stated beliefs on behavior in an innovative way. We elicit first- and second order action beliefs in an incentivized fashion and control for a potential

---

<sup>6</sup>Bartling and Fischbacher (2010) also study responsibility shifting among subjects, but among principals in a dictator game variant.

consensus effect using the first order beliefs of the trustee with respect to the other trustee. Our results are in the range of Bellemare et al. (2010) indicating that consensus effects do play a role, but less in size than the effect of second order beliefs on choices. The *leader-follower treatment* gives us an opportunity to study how robust the effect of beliefs is. Follower trustees are i) asked what they believe leader trustees will choose, ii) informed about the choice of the leaders, iii) choose themselves, iv) asked what they believe trustors will decide, and then v) they are asked what they believe trustors expect trustees to do. We also tackle the problem of potential correlations with guilt sensitivity and second order beliefs – another source of confound as described in Bellemare et al. (2010) – by directly measuring guilt sensitivity using a psychological measure – the Test of Self-Conscious Affect-3 (TOSCA-3) by Tangney et al. (1989) – arguably the most widely used way to elicit sensitivities to guilt and shame.

The remainder of the paper is organized as follows. Section 2 discusses relevant theoretical approaches and derives behavioral predictions. In section 3 we describe the experimental design. Results are presented in section 4 and section 5 concludes.

## 2 A formal model and hypotheses

Subjects were confronted with a one trustor two trustee version of the mini trust game also used by Charness and Dufwenberg (2006). One trustor, we will call her A, can either choose a safe option or trust both trustees (call them B1 and B2) to invest in a, only for her, risky project. Investing in the risky project has a material cost for the trustees, but bears no risk for them. This means, that each of the trustors has then the possibility to reciprocate. The game is presented in Figure 1. The games we investigate experimentally differ on the information set of the second trustee's decision nodes: In the first game B2 does not know how B1 decided, while in the second game B2 knows the decision of B1. The game theoretic solution of both games is the same and can be derived via backward induction. A will choose the safe option while neither B1 nor B2 will choose the risky project.

However, we know from numerous previous experiments that the theoretical result considering only material payoffs / selfish preferences do not give accurate predictions of observed behavior. Hence, we consider three possible alternative explanations for other-regarding behavior that can be relevant within the given context: i) fairness or inequity aversion (see outcome-based models of Fehr and Schmidt (1999) or Bolton and Ockenfels (2000)), ii) guilt aversion (see intentions-based models (Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006))), and iii) conformity models (see Akerlof, 1980, Bernheim, 1994, or Manski, 2000) that consider the social interaction between leader and follower. The predictions from models of guilt aversion and inequity aversion (as well as those from models of purely selfish behavior) would not depend on the sequence and information structure of the decisions of the trustees. Instead, conformity models take into account that the follower is informed about the leader's behavior. They would predict that the conforming peer effect leads to a positive correlation between

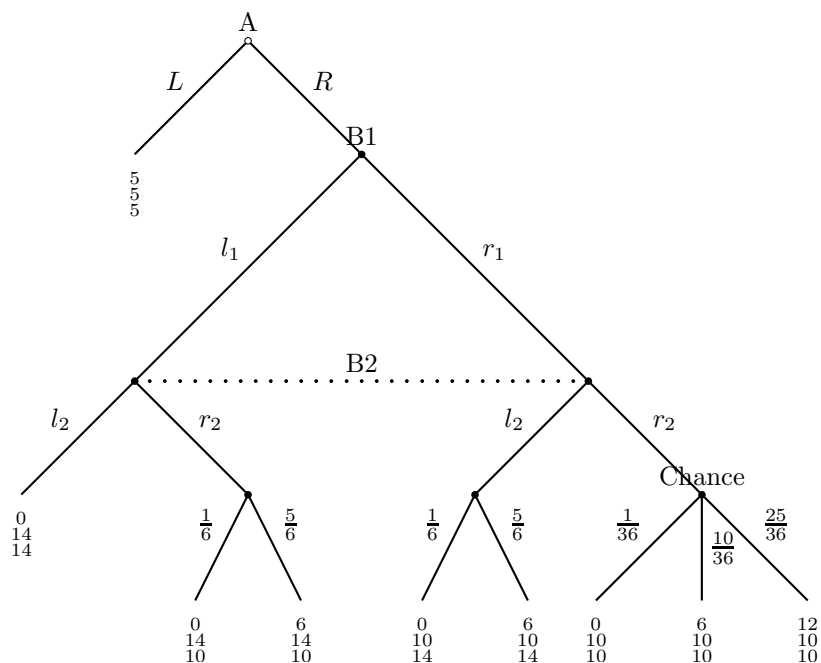


Figure 1: Simultaneous move game – monetary payoffs

leader and follower behavior.

## 2.1 Fairness

We compare the inequality motive between the trustor and the trustee. The utility of the subject depends on the expected payoff inequality. This is a straightforward extension of the model by Fehr and Schmidt (1999). We simplify the model and replace the chance move with the expected outcomes as in this case subjects care only about their expected outcome.

We will assess the game using the utility function introduced in Fehr and Schmidt (1999) under the assumption that inequality in favor is easier to tolerate than adverse inequality. Under this assumption neither in the baseline game, nor in the leader follower game the subjects would engage in the risky project, because the inequality decrease of an investment is too costly to compensate for the losses.<sup>7</sup> Moreover, in both treatments B1 and B2 should behave the same in the same proportions, as the utility calculus they apply is the same for both in equilibrium.

**Hypothesis 1.** *Leader/follower trustees choose to ROLL at the same rate as trustees in the baseline treatment.*

<sup>7</sup>Derivations of the model can be found in the appendix.



## 2.2 Guilt Aversion

Guilt aversion's basic rationale is the following: The more I believe you were disappointed, the more guilt I would anticipate to feel. Hence, the more likely I am to take the kind choice to avoid the negative feeling that would result from the unkind choice. Following Charness and Dufwenberg (2006) individual guilt aversion can be expressed by incorporating  $B$ 's psychological cost of letting down  $A$ . In the two trustee setting of our experiment we slightly adjust  $B$ 's utility to take guilt aversion at a group level into account. Then  $\tau_A^B$  denotes  $A$ 's belief about the probability that  $B$  chooses RIGHT, and  $\tau_B^{AB}$  is  $B$ 's belief regarding  $\tau_A^B$ .<sup>8</sup> In addition,  $\tau_{B1}^{B2}$  denotes  $B1$ 's belief about the probability that  $B2$  picks RIGHT.

In order to measure the amount  $B1$  thinks he hurts  $A$  by his action (choosing LEFT), we calculate the difference between  $A$ 's payoff when  $B1$  plays RIGHT and when he plays LEFT (weighted by the second-order belief  $\tau_{B1}^{AB1}$ ):  $(5 + 5 \cdot (1 - \tau_{B1}^{B2})) \cdot \tau_{B1}^{AB1} - 0$

How much this actually affects  $B1$  is expressed by taking his sensitivity to guilt  $\gamma_{B1}$  into account. Hence, if  $B1$  selects LEFT, he therefore experiences expected guilt of  $(5 + 5 \cdot (1 - \tau_{B1}^{B2})) \cdot \tau_{B1}^{AB1} \cdot \gamma_{B1}$ . This psychological cost of guilt reduces  $B1$ 's material payoff of choosing LEFT. Given  $B1$  is rational she will prefer RIGHT over LEFT if the following inequity holds (note that  $\gamma_{B1} = 0$  represents the model's special case of pure self-interest, when subjects are not affected by negative feelings according to guilt):

$$U_{B1}^{LEFT} = 14 - (5 + 5 \cdot (1 - \tau_{B1}^{B2})) \cdot \tau_{B1}^{AB1} \cdot \gamma_{B1} < 10 = U_{B1}^{RIGHT} \quad (1)$$

This leads to hypotheses 2 and 3 which are about the two input variables of the guilt aversion model, second-order beliefs and the sensitivity to experience guilt.

**Hypothesis 2.** *The higher  $B$ 's second-order belief  $\tau_{B1}^{AB1}$  is, the higher is the probability that  $B1$  will choose RIGHT.*

**Hypothesis 3.** *The higher  $B$ 's sensitivity to guilt  $\gamma_{B1}$  is, the higher is the probability that  $B1$  will choose RIGHT.*

Finally, based on equation 1 we suspect a positive interaction effect between the second-order beliefs and the sensitivity to guilt.

**Hypothesis 4.** *There is a positive interaction between  $B$ 's sensitivity to guilt  $\gamma_{B1}$  and  $B$ 's second-order belief  $\tau_{B1}^{AB1}$ .*

## 2.3 Conformity

In neoclassical economic theory only market interactions are of interest, social interactions are not considered. This lack of realism has been addressed, among others, by Akerlof (1980), or Bernheim (1994) who

---

<sup>8</sup>We refrain from modeling higher order beliefs, as in the experiment we elicit beliefs only to the second order. Furthermore this modeling approach is consistent with previous literature.

developed models of social interaction. Manski (2000) discusses various ways how the actions chosen by other agents affect our own decisions, for instance at the preference level. Gaining status or social-esteem is then achieved by conforming to the observed action of the peer. In our context, followers would take the action of the leader as a signal for the social norm in the situation they are in as well.

Evidence for conforming behavior exists in the lab as well as in the field. Güth et al. (2007) find support in a public goods experiment. Leading by example – letting one group member contribute to the public good before followers do – increases contributions in comparison to the standard voluntary contribution mechanism. Shang and Croson (2006) analyzed how information about a typical contribution to a radio station affects contribution rates. They found that the highest reference amount (\$300) leads to a significantly higher contribution than giving no information about a typical contribution at all.

Our context, two agents can exploit the principal or not (which constitutes a moral dilemma), is different from a public goods game setting (a social dilemma). Yet, the basic mechanism is the same. Followers are told what someone else in the same position did. Hence, it should be interesting to check, whether in our experiment conforming to the leader's choice has a similar effect on the follower's motivation. Since the leader's signal can be unkind or kind behavior we distinguish between these two cases.

**Hypothesis 5.** *Followers choose LEFT more often, if their leader has done so.*

**Hypothesis 6.** *Followers choose RIGHT more often, if their leader has done so.*

### 2.4 Guilt Alleviation

Alternatively, cooperation of the leader trustee may lead to unkind behavior of the follower trustee, if we consider collective guilt as a motivation. Knowing that the other trustee already cooperates may reduce responsibility attribution. Guilt-averse trustees who would play kindly in a 2 person setting (to avoid feeling guilty from disappointing the trustor) may now take the opportunistic choice, if they are informed that the trustor already received something from the leader trustee. The feeling of guilt is crowded out. Hence, if responsibility can be shifted to the leader, guilt is alleviated, and as a consequence the follower trustee does not cooperate.

**Hypothesis 7.** *Followers choose LEFT more often, if their leader has chosen RIGHT.*

## 3 Experiment

### 3.1 Design

Subjects played the game three rounds with perfect stranger matching. In the baseline treatment, the trustor and the trustees made the decisions simultaneously. The trustees' choices only mattered, when the trustor chose to trust and subjects were explicitly told about this. Then we elicited the first order

Stage	Treatment	
	Baseline	Leader-Follower
Observations	124	58
I	Decision Trustor	
	Decision Leader Trustee	
II	Decision both Trustees	FO bel. over leader
	Decision Follower Trustee	
III	First Order Beliefs	
IV	Second order beliefs	
V	Risk aversion assessment	
VI	Questionnaire with guilt sensitivity measures	

Table 1: Experimental design

beliefs of the trustor on trustees behavior and the first order beliefs of the respective other trustee.<sup>9</sup> Then we asked for the second order beliefs of the trustees on trustors' expectations. In the leader follower treatment, the order was the same, except that the leader trustee made the decision before the follower trustee and the decision was reported to the follower trustee *before* she made a decision.

Table 1 summarizes the experimental procedure. Second order beliefs were elicited *after* the decisions were made in order to avoid that subjects think too much about this before and to allow for self justification, which works against our hypothesis that subjects shift guilt, which makes our results on beliefs even more conservative. We deliberately refrained from introducing the possibility to communicate as we believe that communication not only shifts the beliefs but can also trigger other forms of behavioral motivations: it creates a form of identity which can shape behavior as has recently been shown in economic experiments by Chen and Li (2009) or Hargreaves-Heap and Zizzo (2009) or subjects keep their (verbal) promises because they are averse to lying and not to guilt (for experimental evidence in this class of games see Vanberg (2008b)).

In the post-experimental questionnaire we assessed subjects' general dispositions with respect to guilt and shame using the Test of Self-Conscious Affect-3 (TOSCA-3) by Tangney et al. (1989), arguably the most widely used way to elicit sensitivities to guilt and shame. It consists of 16 everyday-life scenarios in which something went wrong. For each situation subjects are presented a list of possible reactions

<sup>9</sup>We elicited the distribution of the beliefs using quadratic scoring rule (see Schotter and Sopher (2007)). The intuition of the quadratic scoring rule was explained to the subjects and they were able to practice. The exact formula was available upon request.

(among them a shame- and a guilt reaction) and for each they are asked to rate how likely they are to react in that way. The TOSCA-3 relies on the self-behavior distinction between shame and guilt. Guilt responses are characterized by regret and negative behavior-evaluations (thinking "I made a mistake", for example), as well as repair action tendencies (like apologizing). Shame responses are characterized by negative self-evaluations (thinking "I am a terrible person") and withdrawal action tendencies (e.g., hiding).

A – well validated – index constructed from responses provides us with a measure of the disposition to guilt of the subjects – which is the empirical correspondent to the guilt aversion parameter  $\gamma$ . The TOSCA-3 also provides us with measures for the sensitivity to shame, and pride. For an application in economics see for example Ong (2010).

### 3.2 Procedure

We ran 9 separate sessions for the two treatments. 198 participants were recruited among students from various disciplines at the local university using the ORSEE software Greiner (2004). The experiment was programmed and conducted with the software z-Tree Fischbacher (2007). Subjects were seated separately and received written instructions (see Appendix C for translated materials). After answering questions privately, participants had to answer a few control questions. The experiment only started when all participants had answered all control questions correctly. The sessions took on average 90 minutes, including reading the instructions, answering control questions and payment. Average earnings were €12.76 with minimum €7.00 and maximum €16.60, including a €2.50, show-up fee.

## 4 Results

In the following we analyze results of the experiment and discuss them in the light of the possible motivations presented in section 2 (fairness concerns, guilt aversion, and conformity).

### 4.1 Trustee behavior

**Result 1** Follower trustees "invest" less often than baseline trustees.

Figure 2 illustrates that follower trustees in the leader follower treatment roll the dice significantly less often than trustees in the baseline treatment (ranksum test, t-test:  $p=0.02$ ). They also roll significantly less often than leader trustees (ranksum test, t-test:  $p\text{-value}=0.10$ ). A finding that partially contradicts hypothesis 1, as we do not find a significant difference in the leaders' behavior and the behavior of the baseline subjects (ranksum test, t-test:  $p\text{-value}=0.58$ ). It is crucial to see that this behavior does not seem to be mediated by a change in the second order belief of the follower trustee compared to the leader trustee. Follower trustees have no diverging second order beliefs from leader trustees (ranksum test,

p-value: 0.51, t-test, p-value: 0.57) and the average second order belief of all trustees lies at around 35%. So if followers would experience guilt concerning their own action, they would choose equally often to roll than in the baseline treatment. This seems to suggest that feelings of guilt are crowded out due to favorable behavior of the leader, even though followers hold – in an average sense – the same beliefs as the leaders (followers’ belief that they should go right is 34% while leaders’ belief is 37%, this difference is not significant, ranksum-test, p-value: 0.507).

**Result 2** Cherry picking: Follower trustees conform with the leader’s choice only when it is to their advantage.

Does the decision of the leader trustee have a conforming effect on the choice of the follower? Table 2 shows the decision of trustees distinguishing between baseline trustees, and follower trustees conditional on their leader’s decision. If the leader trustee chose not to roll, 36 out of 43 or 83.72% followers also chose not to roll. Compared to the baseline (87 out of 124 or 70.16%), follower behavior after the leader chose not to roll is less kind (Wilcoxon ranksum test, p-value: 0.051). It appears that conformity is a relevant motivation if to conform means a material advantage for oneself, supporting hypothesis 5. If the leader trustee chose to roll, 14 out of 15 or 93.33% followers chose not to roll. Follower behavior after the leader chose to roll is clearly not conforming to the leader trustee’s choice, in contrast to hypothesis 6 and even less kind than in the baseline (Wilcoxon ranksum test, p-value: 0.045). When conforming implies a material disadvantage it seems that trustees do not care about following the leader’s choice. Instead, we find that followers reciprocate less often (6%) if the leader did reciprocate, pointing into the direction of the collective guilt hypothesis.

Given those results, we tested whether the cherry picking behavior of followers can be replicated when the followers’ decision is asked in a strategy method design (Selten, 1967). Follower subjects were not informed about the leader’s choice before they had to choose. Instead, they had to make two decisions, one in case the leader had played cooperatively, one assuming the leader did not cooperate. From a game-theoretic perspective this difference in the elicitation method should not make a difference. Table 3 clearly shows that most of the subjects make their choice not dependent on the leader’s choice. Furthermore, this results in a distribution of choices given the leader went LEFT (73.33%) / RIGHT (70%) that nearly equals the distribution of the baseline simultaneous move game (70.16%). Interestingly, when choices are elicited with the strategy method, the strong direct response results disappear. The strategy method is widely used tool in experimental economics and in previous comparisons only little differences in choices have been reported.<sup>10</sup> However, Brosig et al. (2003) find a significant hot (direct response) vs. cold (strategy method) effect in a bargaining game with a punishment option. Our result suggests that the strategy method delivers potentially different results, when the options/counterfactuals cannot easily be ranked in a linear way, as for instance the trustee’s response to increasing transfers of the trustor in a

<sup>10</sup>See, for instance, Casari and Cason (2009) or Rauhut and Winter (2010).

Figure 2: Trustee choses to "ROLL"

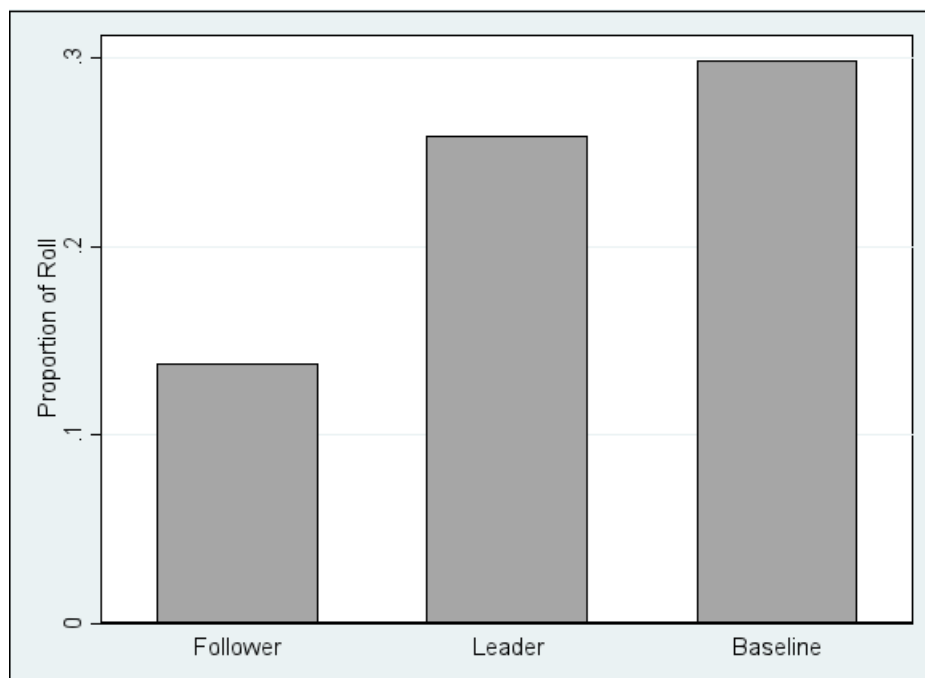


Table 2: Comparison of decision to reciprocate in Baseline vs Leader Follower Treatment

	All	Follower's decision	
	Baseline	Leader left	Leader right
Left	70.16%	83.72%	93.33%
	(87)	(36)	(14)
Right	29.84%	16.28%	6.67%
	(37)	(7)	(1)
p-values	<i>Baseline vs. Leader left</i>		<i>Baseline vs. Leader right</i>
$\chi^2$ test	0.05		0.05
Fisher's exact test	0.06		0.07
One sided Fisher's exact test	0.04		0.03

Number of observations in parenthesis.

*Note:* This table reports the percentage of subjects who reciprocated. The first column shows the results for the baseline treatment, when decisions were made simultaneously. Columns two and three report the followers' decision, depending on the decision of the leader. The test statistics compare the baseline treatment to the followers' decisions conditioning on the leaders' decision.

Table 3: Comparison of decision to reciprocate in the strategy method for followers

Follower type		Conditioning on leaders choice	
Left, uncond.	63.33% (19)	Leader left	73.33% (22)
Right, uncond.	20.00%(6)	Leader right	70.00% (21)
Compensators	6.67% (2)		
Conformist	10.00% (3)		
Total	100% (30)		

*Note:* This table reports the percentage of subjects who reciprocated and the absolute numbers in brackets, given the leader's decision. We categorize the subjects into four mutually exclusive types: (1) Unconditional left: subjects who never reciprocate, (2) Unconditional right: Subjects who always reciprocate, (3) Compensators: Subjects who reciprocate when the leader does not and don't reciprocate when she does (4) Conformists: Subjects who mimic the decision of the leader. The right hand side of the table gives the percentage of reciprocation conditional on the leader's choice in order to provide an easy comparison to the results on the conditional decisions of the followers from table 2.

trust game (assuming the trustee wants to reciprocate). When the underlying decision processes are more complex, the strategy method may dilute behavior. In our context, cherry picking in the strategy method requires to enter two seemingly inconsistent decisions in subsequent screens which may be individually costly for the follower subjects. In the direct response method followers can adhere to cherry picking in a much easier way as only one decision is requested and they do not have to ponder about the other choice and the resulting inconsistencies.

**Result 3** There is a correlation between second order beliefs and actions. It is higher for subjects who have a higher measured guilt sensitivity. This is true for all treatments.

We consider the role of second order beliefs in a regression framework in order to evaluate their significance for behavior. In order to avoid the critique of Vanberg (2008b) on using promises as second order belief shifters, we directly elicit second order beliefs to examine the correlation between them and observed behavior. In table 4 column 1 shows the pure treatment comparison, column 2 adds the behavior of the other trustee, column 3 considers second order beliefs, and column 4 guilt sensitivity.

The first two columns confirm the non-parametric tests that there exists a substantial wedge in behavior between leaders and followers and show that there seems to be a negative correlation between leader and follower behavior, however insignificant.

Column 3 confirms the positive and highly significant correlation of second order beliefs on cooperative behavior of trustees, in line with guilt aversion and found in previous studies. The coefficient of the second order belief is significantly lower for followers (0.430), than for leaders (1.139), however still statistically significantly positive (5%-level). To further examine the role the second order belief takes in describing

behavior we add our measure for guilt sensitivity in column 4 of table 5, as well interacting it with the role of the subjects and the second order beliefs. The main effect of guilt sensitivity on the decision to ROLL is positive and significant at the 5%-level. Most importantly, the three-way interaction between guilt sensitivity, second order beliefs, and followers is positive and significant.

The observed behavior is not predicted by inequity aversion as there is a significant difference in trustee behavior between baseline and leader-follower treatment. It is also at odds with conformity seeking, which would suggest that the choice of leader trustees influences the behavior of follower trustees in a positive way. The resulting behavior is in line with what guilt aversion predicts as second order beliefs and the sensitivity of guilt are both correlated with the kind choice.

However, the choices by follower trustees deviate from this pattern. Their behavior is more consistent with collective guilt aversion where follower trustees shift responsibility to the leader trustee who has already reciprocated. It seems followers – knowing that the trustor will not be empty-handed when they defect – feel less obliged to reciprocate. This reasoning should not disturb the unconditional belief of the follower trustee about the expectation of the trustor, because the follower knows that the trustor does not know how the leader will behave.

Interesting to see in this context is then, whether the leader actually thinks that she has an influence on the follower and if yes, in which direction? A simple regression of the action of the trustees and their first order beliefs on what the other trustee does shows that there is a significant correlation between the leaders' belief over the followers' action (0.719, p-value < 0.01), while there is a positive, but less strong correlation (0.256, p-value: 0.133) between the followers' belief and the leaders' action. The former correlation indicates that the leader thinks she has a positive influence on the follower. The latter hints at a false consensus effect, however the difference between the two coefficients is only marginally significant at a 10% level. This observation leads us now to our strategy to control for the potential false consensus effect and to make causal arguments regarding the first order beliefs.

### 4.2 Controlling for potential false consensus effects

The false consensus effect is a potential reason why subjects' beliefs are correlated with actions Ross (1977) or Engelmann and Strobel (2000). Vanberg (2008a) shows in a simple model that when beliefs of the state of the world are drawn from a common but unknown distribution and the signals, there is a correlation of the beliefs over the action of the other and the belief over his action and higher order beliefs. Within our regression analysis this poses the problem of unobserved variables bias as we do not control for the subjective belief over the state of the world or belief over the distribution of traits in a society, and the cause of the second order belief. However, we claim that it is relatively easy to find a proxy for the unobserved variable by measuring the belief how *other trustees* behave. Asking for this belief appears to be particularly credible in our 3-player design, since there actually exists one other trustee. We claim that the belief over the behavior of the other trustee is the subject's belief over the distribution of the



Table 4: Trustee behavior

	(1)	(2)	(3)	(4)
Follower	-0.171*** (0.064)	-0.181** (0.078)	0.104 (0.075)	1.308** (0.533)
Leader	0.123** (0.050)	0.125** (0.063)	0.025 (0.081)	-0.140 (0.288)
Age	-0.010 (0.017)	-0.011 (0.017)	-0.008 (0.012)	-0.005 (0.013)
female	-0.085 (0.071)	-0.082 (0.071)	-0.075 (0.056)	-0.112** (0.056)
Other Trustee right		-0.109 (0.078)	-0.020 (0.069)	-0.019 (0.070)
Follower × other right		0.020 (0.117)	-0.101 (0.105)	-0.111 (0.105)
Leader × other right		-0.087 (0.167)	0.002 (0.158)	0.009 (0.154)
Second order belief			1.139*** (0.134)	2.882** (1.349)
Follower × s.o. belief			-0.709*** (0.236)	-3.871** (1.625)
Leader × s.o. belief			0.191 (0.202)	-0.859 (0.910)
<b>Guilt sensitivity</b>				0.015** (0.007)
Follower × guilt sens.				-0.020** (0.009)
Leader × guilt sens.				0.002 (0.005)
S.o. belief × guilt sens.				-0.028 (0.022)
Foll. × s.o. belief × guilt sens.				0.052* (0.027)
Lead. × s.o. belief × guilt sens.				0.018 (0.016)
Constant	0.590 (0.394)	0.633 (0.402)	0.143 (0.295)	-0.832* (0.502)
<i>Combined coefficients</i>				
Effect Sec ord. bel.: leader			1.331*** (0.245)	
Effect Sec ord. bel.: follower			0.430** (0.187)	
Effect guilt sens.: leader				0.036** (0.016)
Effect guilt sens.: follower				0.047* (0.026)
Observations	240	240	240	240
$R^2$	0.034	0.047	0.300	0.336
Clusters	144	144	144	144

Standard errors in parentheses

Linear probability model, where the dependent variable is trustees choosing to 'roll'.

Comparisons to the baseline, where trustees decide simultaneously.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

unobserved trait in the world. Controlling for this effect allows us to interpret the second order belief as causal and the estimate of the size of the coefficient becomes more accurate.<sup>11</sup>

The first two columns of table 5 show basic regressions with the first order belief over the other trustee (1), and with the second order belief (2). All observations with available first order belief over the other trustee are used ( $N = 64$ ). Used separately both are highly significant. Column 3 shows that, when controlling for first order beliefs of the other trustee, the coefficient of the second order beliefs drops from 0.842 to 0.566, suggesting subjects are prone to a false consensus effect of roughly the same magnitude as estimated by Bellemare et al. (2010). However, the coefficient on the second order beliefs is still large and significant, which indicates a causal relationship between second order beliefs and decisions, which further supports our claim of the crowding out of guilt.

Column 4 finally shows the importance of guilt sensitivity, measured by the TOSCA-3 scale. While the coefficient of the second order beliefs drops substantially and gets – insignificantly – negative, the interaction of the measure of guilt sensitivity and second order beliefs is significant and positive. This means that the influence of the second order beliefs on behavior are more important for subjects who are more sensitive to guilt, while the main effect measures the effect of the second order beliefs on the least guilt sensitive subject.

We use the similar strategy to measure the belief of the leader on her influence on the follower. We pool the data from the leader-follower treatment with the baseline treatment. Then we regress the first order belief of the B1s on their choice, and interact this with the leader. The B1s in the leader-follower treatment are the leaders, so if we believe that randomization did work, we can interpret the coefficient on the interaction term as the average leader's belief of her influence, having controlled for the false consensus effect in an average sense. This coefficient is positive, however insignificant so the belief over her actual influence is rather low.

### 4.3 Trustor behavior

When we compare the total rolling behavior of the trustees between the two treatments, we find that on average 30% of the trustees roll in the baseline treatment, while – on average over leaders and followers – only 20% do so in the leader follower treatment. This difference is significant at the 10% level (test of proportion, p-value: 0.074). However, this does not seem to be anticipated by trustors. They believe that subjects will roll 35% of the time, irrespective of the treatment (t-test, p-value: 0.464, ranksum test, p-value: 0.581). This non-difference in beliefs reflects into a non-difference in behavior, where around 32% of the trustors choose to trust in both treatments (test of proportion, p-value: 0.740).

Table 7 shows the results of the regression of the choice of the trustor on first and second order

---

<sup>11</sup>A different approach is taken by Bellemare et al. (2010) who draw the identification from informing subjects of the first order beliefs of the principal and from situations where they vary the degree of guilt subjects may feel, including situations where they should not feel guilty at all when taking a particular action.

Table 5: Trustee behavior: Tackling false consensus

	(1)	(2)	(3)	(4)
Follower	-0.137* (0.070)	-0.137** (0.067)	-0.137** (0.067)	0.081 (0.099)
Leader	0.123** (0.061)	0.103* (0.057)	0.114** (0.057)	0.040 (0.099)
Trustee: First ord. belief other	0.749*** (0.122)		0.436*** (0.127)	0.382*** (0.138)
Age	-0.001 (0.015)	-0.002 (0.015)	0.000 (0.014)	0.003 (0.013)
female	-0.074 (0.070)	-0.096 (0.068)	-0.091 (0.066)	-0.115* (0.064)
Second order belief		0.842*** (0.148)	0.566*** (0.166)	-0.798 (0.775)
Other Trustee right				0.023 (0.079)
Follower × other right				-0.167 (0.127)
Leader × other right				0.072 (0.150)
Follower × s.o. belief				-0.463* (0.262)
Leader × s.o. belief				0.124 (0.201)
o.Trustee: Second order belief				0.000 (.)
Guilt sensitivity				-0.000 (0.004)
SO belief × guilt sens.				0.026** (0.013)
Constant	0.082 (0.359)	0.106 (0.362)	-0.009 (0.328)	-0.102 (0.445)
<i>N</i>	64	64	64	64
<i>R</i>	3	3	3	3
Observations	192	192	192	192
<i>R</i> <sup>2</sup>	0.220	0.240	0.284	0.343

Standard errors in parentheses

Linear probability model, where the dependent variable is trustees choosing to 'roll'.

Comparisons to the baseline, where trustees decide simultaneously.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 6: Believed influence of leaders

B1 right	0.143* (1.71)
LF	-0.0754 (-1.35)
Right X LF	0.129 (1.49)
Constant	0.359*** (7.67)
Observations	96

*t* statistics in parentheses

Cluster robust standard errors, clustered by ID

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

beliefs interacted with the treatment variable, similar to the tables for the trustees. We do not observe strong treatment effects when controlling for second order beliefs. Those, however, are strongly correlated with actual behavior, even more than the first order beliefs, which would represent the actual materially relevant variable determining leaders' choice.

## 5 Discussion

Our results show that trustees are in fact driven by guilt aversion: their second order beliefs in connection with their sensitivity to guilt give them orientation about the disappointment they would inflict on the trustor. However, our results also suggest that the positive effect of guilt aversion on cooperative behavior easily vanishes, if stronger – more direct – signals are available to trustees. In our context the kind choice of the leader trustee appears to be such a direct signal. While the processes of guilt aversion (second order beliefs and sensitivity to guilt have a positive effect on behavior) still appear to work for follower trustees, they choose the kind option significantly less often. In particular they do so when the leader has chosen to play kind. It seems that follower trustees tend to *shift responsibility* to the leader, and disregard their second order belief. Shafir et al. (1993) argue that subjects always chose reasons why they act as they do. The leader gives a suggestion of how to behave in a socially ambiguous situation, but interestingly she seems to be perceived in an asymmetric way. If the leader trustee took responsibility, the follower trustee tends to play unkind. It appears the follower is happy to shift the responsibility towards the trustor to the leader trustee. Following example is not attractive when it involves a material disadvantage. Instead, it seems that the leader's action serves to salve the follower's conscience and the path is clear for performing a selfish act. If the leader trustee did not cooperate, the story is different, yet the result is the same. When following example involves material gains, follower trustees seem to be willing to conform to the leader's action and play unkind as well. Interestingly, this kind of cherry picking

Table 7: Trustor behavior

	(1)	(2)	(3)	(4)
LF	-0.0284 (-0.33)	-0.00843 (-0.12)	0.0528 (0.57)	0.0528 (0.57)
Trustor: First order belief		0.396* (1.93)	0.331 (1.56)	0.529* (1.70)
Trustor: First order belief over leader		0.0504 (0.20)	0.183 (0.59)	-0.0150 (-0.05)
Trustor: Second order belief		0.820*** (3.62)	0.829*** (3.57)	0.829*** (3.57)
LF × First order belief over leader			-0.198 (-0.76)	
Risk aversion			-0.0238 (-1.14)	-0.0238 (-1.14)
LF × First order belief				-0.198 (-0.76)
Constant	0.339*** (5.59)	-0.110* (-1.71)	-0.0997 (-1.14)	-0.0997 (-1.14)
Observations	120	120	120	120

*t* statistics in parentheses

Cluster robust standard errors, clustered by ID

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

behavior – following the signal or applying the social norm that provides the best monetary payoff – has also been found in other contexts, e.g. Winter et al. (Forthcoming).

The positive effects of guilt aversion on behavior seem to be easily dominated, when other relevant input for the decision to be made is available. This liability of being replaced by other signals – if there are some – may explain the differing results in existing studies about guilt aversion. The original article of Dufwenberg and Gneezy (2000) but also Charness and Dufwenberg (2006) and others do find a positive correlation between *second order beliefs* and cooperative behavior. However, Ellingsen et al. (2010) and others (Kulisa and Römer (2009) or Kawagoe and Narita (2010)) fail to find such a correlation. These studies use the reported beliefs of trustors to install second order beliefs in trustees. As we have illustrated second order beliefs are prone to be disturbed by other signals, though. They might get crowded out by the slightest doubt about the trustworthiness of these reported beliefs (i.e. trustees may believe they were given strategically), or subjects perceive the reported beliefs as normative.

We also find that a “*false consensus*” effect biases the coefficients of the second order beliefs upward in a simple regression model. We approach this problem by treating the “false consensus effect” as an omitted variables bias, controlling for subjects’ belief of the world by introducing the first order belief over what trustees expect the other trustee does. Our setting, a 3-player design with two trustees, appears to be particularly useful for such an approach since there actually exists one other trustee. Like Bellemare et al. (2010) and Costa-Gomes et al. (2010) who also control for false consensus effects in trust games we still find a substantial effect of second order beliefs. Our procedure reduces the bias roughly by the same amount found in Bellemare et al. (2010). We therefore suggest this as a new and simple strategy to overcome the causality problem of regressions on second order beliefs.

In addition to the role second order beliefs play we also find compelling evidence for the importance of *guilt sensitivity* as a determinant of behavior. The main effect of guilt sensitivity on the decision to ROLL is positive and significant at the 5%-level. Most importantly, the three-way interaction between guilt sensitivity, second order beliefs, and followers is positive and significant. These results stand in stark contrast to the findings by Ellingsen et al. (2010), as the sensitivity to guilt should not have an influence, even in the presence of a false consensus effect.

Finally, it is interesting to see that the theory of inequity aversion has no predictive power in this situation. Even slight modifications, namely from a two player to a three player trust game, make models that are standard in economics by now not explain the relationship of the leader and the follower.

## References

- Bartling, B., Fischbacher, U., 2010. Shifting the blame: On delegation and responsibility.
- Bellemare, C., Sebald, A., Strobel, M., 2010. Measuring the willingness to pay to avoid guilt: Estimation using equilibrium and stated belief models. *Journal of Applied Econometrics* 1, 1.

- Bernheim, B. D., October 1994. A Theory of Conformity. *Journal of Political Economy* 102 (5), 841–77.
- Bolton, G., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90 (1), 166–193.
- Brosig, J., Weimann, J., Yang, C.-L., 2003. The hot versus cold effect in a simple bargaining experiment. *Experimental Economics* 6, 7590.
- Casari, M., Cason, T. N., 2009. The strategy method lowers measured trustworthy behavior. *Economics Letters* 103, 157159.
- Cassar, A., Rigdon, M., 2010. Trust and trustworthiness in networked exchange. *Games and Economic Behavior*.
- Charness, G., 2000. Responsibility and effort in an experimental labor market. *Journal of Economic Behavior & Organization* 42 (3), 375–384.
- Charness, G., Dufwenberg, M., 2006. Promises and Partnership. *Econometrica* 74 (6), 1579–1601.
- Chen, Y., Li, S. X., 2009. Group identity and social preferences. *American Economic Review* 99 (1), 431–457.
- Costa-Gomes, M. A., Huck, S., Weizsäcker, G., 2010. Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect.
- Dufwenberg, M., Gneezy, U., 2000. Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior* 30, 163182.
- Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47, 268–298.
- Ellingsen, T., Johannesson, M., Tjøtta, S., Torsvik, G., 2010. Testing guilt aversion. *Games and Economic Behavior* 68 (1), 95–107.
- Engelmann, D., Strobel, M., 2000. The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics* 3 (3), 241–260.
- Falk, A., Fischbacher, U., 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2), 293 – 315.
- Fehr, E., Schmidt, K. M., August 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114 (3), 817–868(52).
- Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171–178.

- Fischbacher, U., Gaechter, S., 2010. Social preferences, beliefs and the dynamics of free riding in public goods experiments. *American Economic Review* 100(1), 541–556.
- Greene, W., 1993. *Econometric Analysis*. Macmillan, p. 245246.
- Greiner, B., 2004. An Online Recruitment System for Economic Experiments. *GWDG Bericht* 63, 79–93.
- Güth, W., Levati, M., Sutter, M., Van Der Heijden, E., 2007. Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics* 91 (5-6), 1023–1042.
- Hargreaves-Heap, S. P., Zizzo, D. J., 2009. The value of groups. *American Economic Review* 99 (1), 295–323.
- Hermalin, B. E., 1998. Toward an economic theory of leadership: Leading by example. *American Economic Review* 88 (5), 1188–1206.
- Kawagoe, T., Narita, Y., 2010. Guilt aversion revisited: An experimental test of a new model.
- Knez, M. J., Camerer, C. F., 1995. Outside options and social comparison in three-player ultimatum game experiments. *Games and Economic Behavior* 10 (1), 65 – 94.
- Kulisa, J., Römer, D., 2009. Testing guilt aversion in a dictator game, mimeo.
- Latané, B., Darley, J., 1968. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology* 8 (4), 377–383.
- Manski, C. F., 2000. Economic analysis of social interactions. *The Journal of Economic Perspectives* 14 (3), 115–136.
- Ong, D., 2010. Sorting with shame in the laboratory. In: *Proceedings of the Behavioral and Quantitative Game Theory: Conference on Future Directions*. BQGT '10. ACM, New York, NY, USA, pp. 85:1–85:4.
- Potters, J., Sefton, M., Vesterlund, L., 2007. Leading-by-example and signaling in voluntary contribution games: an experimental study. *Economic Theory* 33, 169–182, 10.1007/s00199-006-0186-3.
- Rabin, M., Dec. 1993. Incorporating Fairness into Game Theory and Economics. *The American Economic Review* 83 (5), 1281–1302.
- Rauhut, H., Winter, F., 2010. A sociological perspective on measuring social norms by means of strategy method experiments. *Social Science Research* forthcoming.
- Reinstein, D., Riener, G., Forthcoming. Reputation and influence in charitable giving experiments. *Theory and Decision*.



- Ross, L., 1977. The "False Consensus Effect": An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Experimental Social Psychology* 13 (3), 279–301.
- Schotter, A., Sopher, B., 2007. Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior* 58 (2), 365–393.
- Sebald, A., 2010. Attribution and reciprocity. *Games and Economic Behavior* 68 (1), 339 – 352.
- Selten, R., 1967. Beitrge zur experimentellen Wirtschaftsforschung. J.C.B. Mohr, Tbingen, Ch. Beitrge zur experimentellen Wirtschaftsforschung Die Strategiemethode zur Erforschungdes eingeschrnkt rationalen Verhaltens im Rahmen eines Oligopolexperiments.
- Shafir, E., Simonson, I., Tversky, A., 1993. Reason-based choice. *Cognition* 49 (1-2), 11 – 36.
- Stouffer, S. A., 1949. An analysis of conflicting social norms. *American Sociological Review* 14 (6), pp. 707–717.
- Tangney, J. P., Wagner, P., Gramzow, R., 1989. The Test of Self-Conscious Affect (TOSCA). George Mason University, Fairfax, VA.
- Vanberg, C., May 2008a. A short note on the rationality of the false consensus effect, max Planck Institute Jena.
- Vanberg, C., 2008b. Why Do People Keep Their Promises? An Experimental Test of Two Explanations 1. *Econometrica* 76 (6), 1467–1480.
- Winter, F., Rauhut, H., Helbing, D., Forthcoming. How norms can generate conflict: An experiment on the failure of cooperative micro-motives on the macro-level. *Social Forces*.

## A Randomization checks

<b>Baseline</b>	<i>Age</i>	<i>Female</i>	<i>Risk aversion</i>	<i>Guilt (TOSCA)</i>	<i>Shame (TOSCA)</i>
Mean	22.87	0.64	1.33	60.71	44.54
Std. dev.	2.37	0.48	1.64	6.83	8.75
Minimum	19.00	0.00	0.00	43.00	25.00
Maximum	31.00	1.00	5.00	74.00	63.00
<b>Leader Follower</b>					
Mean	22.03	0.64	1.17	60.96	45.46
Std. dev.	2.40	0.48	1.73	7.96	8.92
Minimum	14.00	0.00	0.00	38.00	19.00
Maximum	28.00	1.00	5.00	75.00	62.00
<b>Difference</b>	0.841*	0.003	0.167	-0.250	-0.917
Observations	168				

Age different because of one outlier, caused by wrong entry of the age (there were no 14 year-old in the sample)

## B Derivations of Predictions

### B.1 Fairness

We analyze the game depicted in 3, however first for the case where all players have full information: the leader follower case. We derive the backward induction solution of the game.

For the decision node of B2 after B1 has chosen  $r_1$ , B2 will chose  $r_2$  if  $10 > 14 - 6.5\beta$ . This implies that  $\beta$  has to be larger than  $\frac{8}{13}$ . So B2 will chose  $r_2$ .

For the decision node of B2 after B1 has chosen  $l_1$ , B2 will chose  $r_2$  if  $10 - 2\alpha - 2.5\beta > 14 - 7\beta$ . Which implies that  $\beta > \frac{8}{9} + \frac{4}{9}\alpha$  which implies that even for an  $\alpha = 0$  even less subjects should chose  $r_2$ .

Given that in both cases B2 choses  $l_2$ , B1 will chose  $l_1$  out of the same reason that B2 choses  $l_2$  on her first decision node. And consequently A shall chose  $L$ . This leaves us with the clear prediction that leaders and followers shall behave the same for the same reason and therefore we should not see a difference in their behavior.

## Experimental Instructions

Welcome and thanks for participating in this experiment.

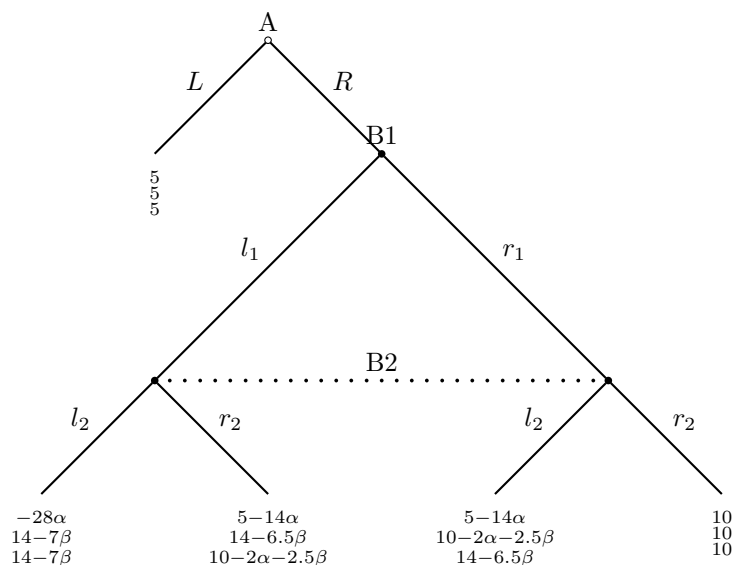


Figure 3: Simultaneous move game – Fehr Schmidt payoffs

In this experiment you can earn a certain amount of money, which depends on your and the other participants' decisions in this experiment. **It is, therefore, very important that you read the following instructions carefully.**

Please be aware that you are not allowed to talk to other participants during the entire experiment. If you have any questions or concerns, please raise your hand. We will answer your questions individually. Please do not ask your question(s) aloud. It is very important that you follow these rules, since otherwise we have to stop the experiment. Please also turn off your mobile phones now.

### General Procedure

The experiment lasts about 90 minutes. Each decision will be explained again briefly on the monitor.

Your payoff from this experiment depends on your decisions and, possibly, the ones other participants make. The exact procedure your payoff is calculated is explained further below.

Your payoff will be calculated in ECU (Experimental Currency Units), 1 ECU = 0,5 EURO. At the end of the experiment your earnings will be converted into EURO and you are paid in cash.

In addition you receive 2,50 Euro as a show-up fee.

After you filled in a questionnaire the experiment ends and you will receive your payoff.

Again the procedure as an overview:

1. Reading of the instructions, test questions (online)

2. Decision situations
3. Questionnaire
4. Payoff and end of the experiment

## Details of the experiment

Always three participants interact with each other. They are called participant A, participant B1 and B2. The following decision situation will be **played several times**, i.e., there will be several periods. Whether you are participant A, B1 or B2 will be determined at the beginning of each period. Hence, it is very important that you **familiarize yourself with each of the roles**. You will only learn about your payoff at the very end of the experiment, not after each period. Out of all the periods that you play, two periods will be relevant for your final payoff.

## Decision Situations

In this game participant A will make a decision first. He/She can decide in favor of option “left” or “right”.

- The choice of “left” implicates a specific payoff, which is 5 ECU for participant A, 5 ECU for participant B1 and 5 ECU for participant B2.
- If participant A chooses option “right”, the payoffs for each of the three participants will be determined by participants B1 and B2.

Then, first B1 and afterwards B2 can choose between two options:

- A decision of “left” means a payoff of 0 for participant A and a payoff of 14 ECU for participant B1 as well as B2.
- A decision of “right” means, that:
  - with a probability of  $1/6$  (approximately 17%) there will be a payoff of 0 ECU for participant A and a payoff of 10 ECU for participant B1 as well as B2.
  - with a probability of  $5/6$  (approximately 83%) there will be a payoff of 6 ECU for participant A and a payoff of 10 ECU for participant B1 as well as B2.

Both, participant B1 and B2, will always be asked for his/her decision, regardless if participant A has chosen “left” or “right”. **None of the participants can observe or will be informed about the decisions the other participants make.**

### Examples for the payoffs:

- If A chooses 'left', all 3 participants will receive 5 ECU from this period.
- If all participants choose 'right' ...
  - it is very likely, that A receives 12 ECU.
  - B1 and B2 each receive 10 ECU.

**That means, A can receive a maximum payoff of 12 ECU, because he/she plays with B1 as well as with B2.**

The following diagrams illustrate the game and the resulting payoffs:

[diagram illustrating the interaction]

**As B2 you will learn about the decision of B1 before you have to make your own decision.**

### **Estimates**

Besides the choice of your actions you will be asked for

- your expectation concerning the other participant's action
- your estimates of the expectation of the other participant concerning your own action

You can earn money with these estimates. **The closer you are to the real amount the more you earn.** Therefore it is important for you to read the instructions carefully.

You are able to split your estimate in different intervals. Please indicate the estimated probability with a value between 0 and 100. Please consider that all probabilities must sum up to 100.

### **Example for stating your estimates**

100 Students took an exam. The possible grades for the evaluation of the exams are 1, 2, 3, 4, and 5.

**You know** that the knowledge about the exam's topic is equally distributed amongst the students who have taken the exam, i.e., there are just as many very good results as there are very poor results. Hence, **every grade is equally probable.**

To maximize your profit your estimate should look like this.

[table with the distribution]

You insert the same probability in each box. There are five intervals, hence, the probability results in  $100 / 5 = 20$ .

## Payoff

Your payoff from these estimates depends on how close your estimates were to the actual results from this experiment. The closer you are to the real amount the more you earn.

**The maximum earning per estimate is 3 ECU.** The actual figure is defined taking all participants into account.

In any case it is best for you to state your actual/true estimates. Upon request you can (after the experiment) look at how your earnings from the estimates were calculated in detail.

## Your payoff from the experiment

You will be paid your earnings in cash directly after the end of the experiment, that means after you completed a final questionnaire.